Low-Power Image Recognition Challenge

Kent Gauen, Rohit Rangan, Anup Mohan, Yung-Hsiang Lu
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana
{gauenk,rrangan, mohan11, yunglu}@purdue.edu

Wei Liu, Alexander C. Berg
Department of Computer Science
University of North Carolina at Chapel Hill
{wliu, aberg}@cs.unc.edu

***Abstract*—Significant progress has been made in recent years using computer programs recognizing objects in images. Meanwhile, many cameras are embedded in battery-powered systems (such as mobile phones, wearable devices, and drones) and energy efficiency is essential. Even though many research papers have been published on the topics related to low power and image recognition, there does not exist a common metric for comparing different solutions in terms of (1) energy efficiency and (2) accuracy in recognition. Low-Power Image Recognition Challenge (LPIRC) is, to our knowledge, the only on-site competition that considers both energy consumption and recognition accuracy. LPIRC was held as one-day workshops in the Design Automation Conference in 2015 and 2016. Each participating team brought their own system to the workshops. The referee system of LPIRC includes (1) an intranet, (2) a power meter, and (3) an HTTP server that provided the images and accepted the answers from the contestants' systems. The scores were the ratio of recognition accuracy and the energy consumption. The winner of 2016 was able to analyze 7,347 images and achieve 9.44% normalized mAP (mean average precision) with average power consumption of 4.7 W. Another team analyzed 1,020 images and achieved 25.7% normalized mAP.**

I. Introduction

Using machines to recognize objects in visual data (image or video) has been a goal for researchers and novelists. Nearly half a century ago, *HAL* in *Space Odyssey* was able to read the lips of Bowman and Poole; this was one of the early depictions of computer vision. Image recognition is a well studied topic; IEEExplore returns more than 26,000 papers when searching "image recognition" in abstracts. Over these decades, significant progress has been made. For example, face detection is available in digital cameras and social networks can frequently identify friends in photographs [1]. Recently, deep learning has been widely used for image recognition [2]. We also see a rise in cameras embedded in many mobile systems, such as wireless phones, wearable systems, and unmanned aerial vehicles. These systems must carry energy (battery or fuel) and energy efficiency is very important. Despite the importance, there is no widely accepted benchmark and methodology comparing different solutions for low-power image recognition.

The Low-Power Image Recognition Challenge (LPIRC) was previously held as part of the Design Automation Conference in 2015 and 2016. It evaluated image recognition systems based on both accuracy and energy consumption. LPIRC used the object detection dataset from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [3] with additional test data specially added for LPIRC. Each image in the test set included one or several objects that belonged to the 200 pre-defined categories. Each participant had 10 minutes, during which the participant's system requested images and returned the categories and bounding boxes. The score for each participant was the ratio of the detection accuracy and the energy consumption [4]. In 2016, 10,000 images were in the test set. The winner recognized objects in 7,347 images with 9.44% mAP and the average power consumption of 4.7 W. Another team analyzed 1,020 images and achieved 25.7% normalized mAP. For comparison, the ILSVRC has no time or energy restriction; the winner of 2016 ILSVRC was able to achieve 66.3% accuracy.

II. Benchmarks

Benchmarks play a crucial role when comparing different ideas. A simple example can be seen in a mile long race. The benchmark is the length of the race (1609 meters) and the time to finish is the reference. Running a mile quickly perhaps provides evidence that one's method of training is superior to others'. An ideal benchmark relies only on a few controllable factors.

LPIRC considers both total energy consumption and object detection accuracy. The energy consumption can simply be measured by the circuit board. For image processing, there exist benchmarks in the form of datasets and objective functions. LPIRC uses the direct measurement of the total energy consumption and mean average precision with a threshold of 0.5 on the intersection over union (IoU) of the bounding box. This section

978-1-5090-1558-0/17/$31.00 ©2017 IEEE

provides the detailed description and justification of the evaluation metrics used in LPIRC.

1) Datasets: A goal in machine learning is to build a single, unified model for all artificial intelligence. tasks [5]. Since the technology has not arrived, machine learning applications are trained for specialized tasks like speech processing, image processing, semantic parsing, etc. Only recently multiple image tasks (localization, detection, and classification) have become incorporated into one model. At 5^{th} in the ILSVRC 2013 object detection competiton, the model called “Overfeat” from New York University was proposed. It was one of the first successful convolutional neural networks to train end-to-end to simultaneously classify, locate, and detect objects [6]. Many specialized datasets are created. For example, semantic processing comes equipped with datasets like PARASEMPRE [7].

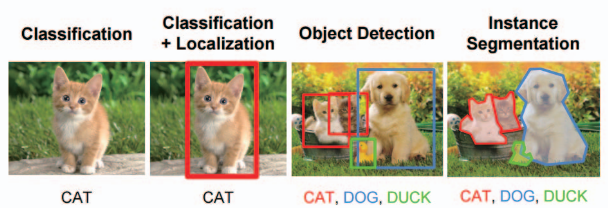
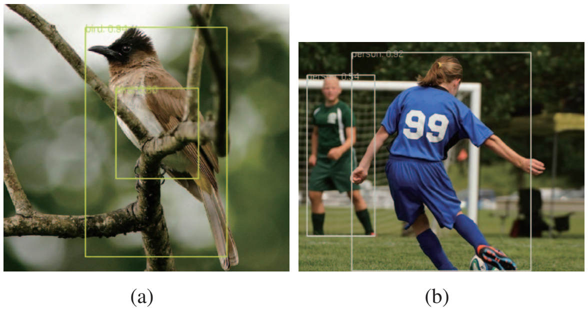


Fig. 1: These figures outlines the definition of many common image tasks [8].

LPIRC considers *object detection*. This means that one must classify and localize each recognizable object in an image (see Figure 1). The figure further distinguishes object detection from other goals in image processing. Classification answers the question: Does this image contain an object that belongs to class x or not? For multi-class classification, it answers: Does this image contain an object that belongs to class x, y, or z? In the figure, it shows that a cat is in the image (far left). Localization is the task of answering: What is the best “bounding box” to fit around object x in the image? The figures shows the bounding box around the cat (middle left). Object detection is classification and localization of *multiple* objects within a single image. This is shown in the figure by the bounding boxes for both cats, the dog, and the duck (middle right). Instance segmentation (far right) is pixel-wise classification. Every pixel is assigned to one of multiple classes, where one of the classes must be “background”. The background is left unlabeled in the figure.

For object detection, there exists a variety of datasets such as: PASCAL VOC, ImageNet, ILSVRC, and COCO [9][10][3][11]. LPIRC uses the ILSVRC dataset, since it is the largest available dataset and is well-known in the machine learning community. The test set for LPIRC is then a subset of the ILSVRC dataset with some new images added specially for LPIRC and labeled with the same labeling methodology. Examples are shown below in Figure 2.



(a)

(b)

Fig. 2: Two samples in ILSVRC-2015 dataset [3].

2) Evaluation Metric: LPIRC uses mean Average Precision (mAP) to measure the accuracy of object detection methods, following ILSVRC [3]. A method produces arbitrary number of detection results for each object classes in each image. Each detection result has the format of (b_{ij}, s_{ij}) for image I_i and object class C_j, where b_{ij} is the bounding box and s_{ij} is the score.

For bounding box evaluation, LPIRC uses intersection over union (IoU). In Equation (1), x is the reported bounding box region and y is the ground-truth bounding box region. They are both identified by two pairs of pixel coordinates.

Intersection Over Union:

$$IoU = \frac{x \cap y}{x \cup y} \tag{1}$$

To accommodate small objects, we loose the threshold using Equation (2), which in practice only affect any objects which are smaller than approximately 25 × 25 pixels [3]. In Equation (2), the 10 comes from giving a 5-pixel margin on each side of the image, which is the average human annotation error according to [3].

$$\mathrm{thr}(B) = \min\left(0.5\ ,\ \frac{wh}{(w+10)(h+10)}\right) \tag{2}$$

The detection results are first sorted in descending order based on detection scores, and are then greedily matched to the ground truth boxes. A detection result is considered as a true positive if the intersection over union (IoU) overlap with a ground truth box is more than the threshold as defined in Equation (2); otherwise it is considered as a false positive.

Note that we also penalize duplicate detections. In other words, if there are multiple detections for an object (e.g. IoU threshold > 0.5), only the detection with highest score is a true positive and all others are false positives. Given this information, we can then compute precision as the fraction correct detections among all the detections reported, and recall as the fraction of detected ground truth objects, for each object class. We then compute average precision (AP) as the area under the precision/recall curve for each object class, and mAP as the average from all object classes.

There are many known methods to measure energy consumption. LPIRC uses the Yokogawa WT310 Digital Power Meter, and the scores are reported in watt-hours. Our final evaluation metric is the mean average precision divided by the total energy consumption in the 10 minute interval. The final score is given by:

$$\text{Total Score} = \frac{\text{mAP}}{\text{Total Energy Consumption}} \tag{3}$$

III. Low-Power Image Recognition Challenge

Low-Power Image Recognition Challenge (LPIRC) compares different systems' ability to recognize objects in images while using as little energy as possible.

A. Design Principles

Figure 3 shows the architecture of the system. The organizer provides a referee system (a laptop), a power meter, and a router. A contestant's system can connect to the router using an RJ-45 cable or Wifi. To prevent interference, this router is not connected to the Internet. Also, the router is protected by passwords and only one contestant's system is allowed to connect to the router at any moment.

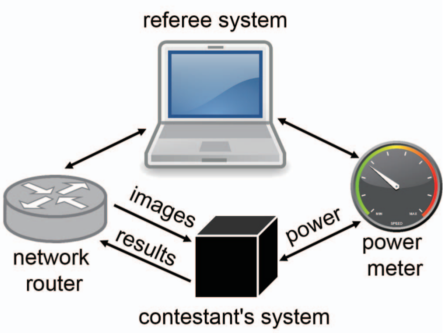


Fig. 3: The system architecture of LPIRC.

LPIRC is an on-site competition: each participating team brings their own system. Designing LPIRC follows these principles:

- Impose as few restrictions as possible. The participants may bring any system and it may consume AC or DC. To accommodate the wide range of possibilities, a Yokogawa WT310 Digital Power Meter is used because it can measure both AC and DC. Also, WT 310 accepts a wide range of power consumption. It can measure current between 5mA and 20A and voltage between 50mV and 10V (DC) and up to 240V (AC). WT 310 is programmable: a program can start and stop a measurement and WT 310 automatically calculates the cumulative energy consumption.
- Make the software as flexible as possible. The referee system and the contestant's system need to communicate so that images can be sent to contestant's system and the recognition results can be sent back to the referee system. LPIRC uses HTTP for data exchange because HTTP is supported by many programming languages and operating systems. The contestant's system uses the GET command to retrieve images and the POST command to send results.
- Allow asynchronous communication. Each GET command must include the image's ID (starting from one). Each POST command also includes the image's ID. This allows the contestant's system to perform recognition while communicating with the referee system in parallel.

The following is the procedure for a contestant's system:

1) Before the competition: Download the training images and the referee source code from http://lpirc.net.
2) Create a system that can recognize images similar to the training dataset and communicate with the referee system.
3) During the competition, the contestant's system logs into the referee system. The referee starts a timer of ten minutes and resets the power meter.
4) The contestant's system uses HTTP GET to obtain images and uses HTTP POST to send answers. Each answer includes six numbers: the image's ID, the category, and four numbers as the bounding box of the recognized object.
5) Within ten minutes, the contestant's system can log out to stop the power meter. Otherwise, when the timer expires, the referee also stops the power meter.
6) The referee system calculates the final score as described in the benchmark section. It is the ratio

of the recognition accuracy and the cumulative energy consumption. An accurate recognition must have the correct category and the reported bounding box must overlap with the ground truth by at least 50%.

B. Input by Camera

In 2016, a separate track provides images through a display, as shown in Figure 4 (a). A contestant's system obtains the images through a camera, and the purpose of the camera is to simulate a human eye. While the pictures were originally taken with a camera, many of the images are high-quality. It is unlikely that practical settings would allow for such ideal photographs. We are interested in how well a camera performs in real-time, as compared to the carefully taken and selected photos given in most datasets.

The referee system changes the images under the requests of the contestant's system. The margins of each image encodes the image's ID, the width, and the height, as shown in Figure 4 (b). If an image's size is greater than the number of pixels available in the display, the image is re-sized while maintaining the aspect ratio.

An example of the image input from camera can be seen in Figure 5. Figure 5 (a) shows the image from the computer to the monitor. Figure 5 (b) shows the image acquired by the camera into the participant's machine from the monitor. The glare is not on the original image. This kind of noise in the image is the goal of using a camera for the input—it is more realistic.

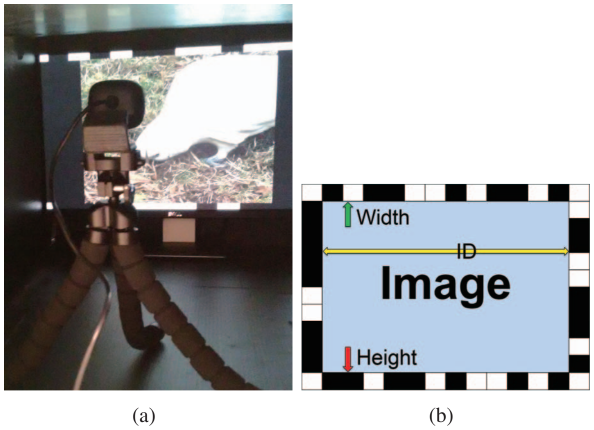


Fig. 4: (a) Use a camera to obtain input images. (b) Format of an image to be captured by a camera.

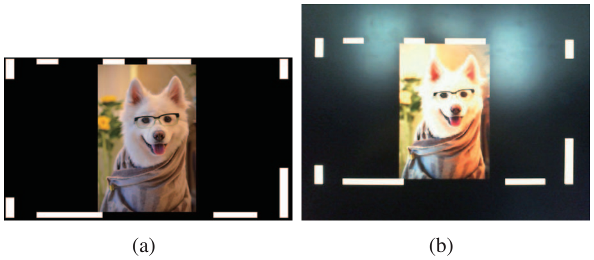


Fig. 5: (a) The input image from the computer to the monitor (b) The output image from the monitor through the camera.

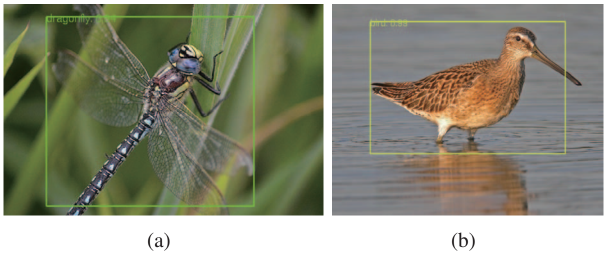


Fig. 6: Two images correctly recognized by a team in 2016. (a) Dragonfly. (b) Bird.

C. Sample Images

Figure 6 shows two images that are correctly recognized by a team in 2016. Figure 6 (a) is a dragonfly and Figure 6 (b) is a bird. As can be seen in these examples, a correct recognition must have the correct category and the bounding box must also be correct.

Figure 7 shows two images that are incorrectly recognized by a team in 2016. In Figure 7 (a), the dogs are recognized correctly but the bounding boxes are wrong. In Figure 7 (b), it is a bird but the team's program marks it as a person. Moreover, the bounding box is also wrong.

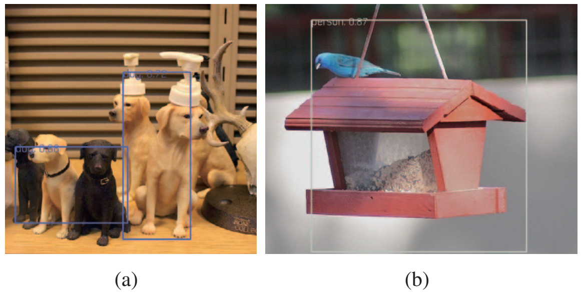


Fig. 7: Two images incorrectly recognized by a team in 2016.

D. Past Results

Results from previous competitions have been recorded and analyzed. Table I contains a summary of results from 2016 and 2015. The average score improved by 5.8 times ($\frac{0.02280}{0.00393}$). The average score of the top 5 teams improved by 2.4 times. The details of each year's results can be viewed in Table II for 2016, and III for 2015. A scatter plot of mAP versus energy consumption of both tables is given in Figure 8. Three teams from 2015 are not displayed because their power consumptions are significantly higher than the other teams' power consumptions. In Table III, only the top-10 scores were kept. Figure 9 shows that mAP scores and reported objects-per-second are correlated.

In LPIRC, the accuracy per image is not the most important indicator of a high rank. A team would not win if they could perfectly recognize the objects in only one image. Instead, a winning team must be able to recognize many images correctly. The mAP is calculated based all test images. In Table II, rank 6 gives the highest normalized accuracy of 25.7% mAP, while the winner's normalized accuracy was only 9.44% mAP. The only normalized accuracy lower than the winner was rank 7's 6.66% mAP, and rank 7 was the only team using the camera as input. The normalized accuracy is caluculated as:

$$\text{Normalized Accuracy} = \text{mAP} \times \frac{20,000}{\text{\# of images processes}}$$

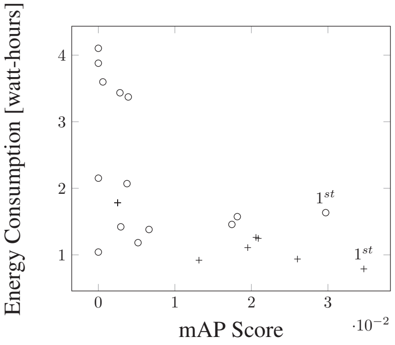


Fig. 8: LPIRC 2016 ("+") and 2015 ("o") mAP versus power consumption results.

2015 & 2016 Summary		
Year	All Score Mean	Top-5 Score Mean
2016	0.02280	.0245
2015	0.00393	.0102
Improvement	5.80215	2.4020

TABLE I: Summary of results from LPIRC 2016 and 2015

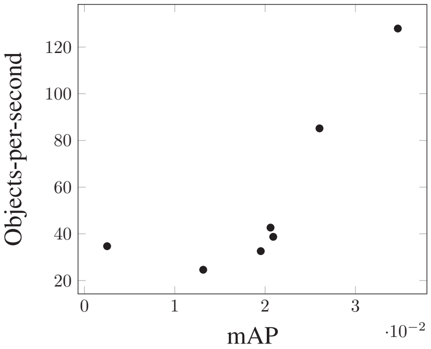


Fig. 9: 2016 mAP versus objects-per-second reported.

2016 RESULTS				
Rank	mAP	Energy	Score	# Images
1	0.03469	0.7891	0.04369	7347
2	0.02602	0.9371	0.02777	4900
3	0.01952	1.1095	0.01760	1779
4	0.02905	1.2475	0.01676	1824
5	0.02060	1.2625	0.01632	1758
6	0.01315	0.9199	0.01430	1020
7*	0.00251	1.7835	0.00141	755

TABLE II: LPIRC 2016 contestant rankings and results breakdown. Energy is measured in watt-hour. *Note that team 7 results are from using the camera as input.

2015 RESULTS			
Rank	mAP	Energy	Score
1	0.02971	1.634	0.01818
2	0.01745	1.457	0.01198
3	0.01816	1.574	0.01540
4	0.00662	1.381	0.00479
5	0.00519	1.183	0.00439
6	0.00294	1.421	0.00207
7	0.03222	16.509	0.00195
8	0.00374	2.070	0.00181
9	0.00391	3.372	0.00116
10	0.00283	3.435	0.00082

TABLE III: LPIRC 2015 contestant rankings and results breakdown. Energy is measured in watt-hour.

IV. Conclusion

LPIRC is the only known competition to consider both accuracy of object detection and power consumption. There exist many possible uses of low-power devices which can detect objects in images. This competition encourages innovation in the field. By using mAP and power consumption, LPIRC established a benchmark for low-power image detection. In the first two years, there are already large improvements in both mAP and energy.

Acknowledgments

Low-Power Image Recognition Challenge (LPIRC) is part of the IEEE Rebooting Computing Initiative. LPIRC's sponsors include IEEE Council on superconductivity, IEEE Council on Electronic Design Automation, Nvidia, and Xilinx. The authors want to express special gratitude to Elie Track, Bichlien Hoang, David Mountain, and David Kirk.

This project is supported in part by NSF ACI-1535108. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Erik Hjelmas and Boon Kee Lows. "Face Detection: A Survey". In: *Computer Vision and Image Understanding* 83.3 (2001), pp. 236–274.

[2] Yoshua Bengio. "Learning Deep Architectures for AI". In: *Foundations and Trends in Machine Learning* 2.1 (Jan. 2009), pp. 1–127. ISSN: 1935-8237.

[3] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252.

[4] Y. H. Lu et al. "Rebooting Computing and Low-Power Image Recognition Challenge". In: *IEEE/ACM International Conference on Computer-Aided Design*. 2015, pp. 927–932.

[5] P. Domingos. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books, 2015. ISBN: 9780465065707.

[6] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks". In: *CoRR* abs/1312.6229 (2013).

[7] J. Berant, A. Chou, R. Frostig, and P. Liang. "Semantic Parsing on Freebase from Question-Answer Pairs". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2013.

[8] Fei-Fei Li, Andrej Karpahy, and Justin Johnson. "CS 231: Convolutional Neural Networks for Visual Recognition". 2016.

[9] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. "The Pascal Visual Object Classes Challenge: A Retrospective". In: *International Journal of Computer Vision* 111.1 (Jan. 2015), pp. 98–136.

[10] Fei-Fei Li, Kai Li, Olga Russakovsky, Jonathan Krause, Jia Deng, and Alex Berg. *ImageNet*. http://image-net.org/.

[11] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft COCO: Common Objects in Context". In: *CoRR* abs/1405.0312 (2014).